

# JUAN VÁZQUEZ MONTEJO | DATA ENGINEER

Oakland CA | Cel: +52 9991260083 | juanvazqmont@gmail.com | github+twitter: @juanvazqmont

---

## EDUCATION

**Physical engineering, 7th may 2003. GPA: 3.0/4 Engineering Faculty.** Universidad Autónoma De Yucatán. **Numerical statistical mechanics thesis:** Simulation time calculation and improvement of an hybrid algorithm Metropolis-Q2R (code generated in **C** and scripts automated in **shell bash**). **Relevant coursework:** Quantum Mechanics, Statistical Mechanics, Mathematical methods.

**Masters degree on physics, 8th july 2005. GPA: 3.7/4 Centro de Investigación y Estudios Avanzados, I.P.N.** Campus Zacatenco with an **Econophysics thesis using time series correlations:** Analysis of exogenous processes in stock market prices. Code simulated in **Fortran 77** and scripts automated in **bash**, thesis document in  $\text{\LaTeX} 2_{\epsilon}$ . **Relevant coursework:** Econophysics, Quantum field theory, Electrodynamics and Statistical Mechanics.

**Physics PhD. 24 august 2012. GPA: 3.7/4 Centro de Investigación y Estudios Avanzados, I.P.N.** Campus Mérida. with a **Complex networks thesis:** Economic exchange in complex networks. Agent-based model in a scale-free Barabasi network. Simulation generated in **C** language and automated in **bash shell**, thesis document in  $\text{\LaTeX} 2_{\epsilon}$ . **Relevant coursework:** Econophysics, Network science, Critical phenomena and complexity.

## PROFESIONAL EXPERIENCE

### Data analyst Plenumsoft. México 8/2016 - 9/2017

**Worked on overdemand prediction in a supply-chain (bullwhip effect).** I adapted algorithms in **R** by analyzing the logfiles of all history available of previous years. The algorithm worked with time-series and dynamically decides if the query is real or not with a 85 percent of accuracy, that means passing from 12 out of 20, to 17 out of 20 real positives in trials.

**Designed a clustering algorithm.** to group users of a power grid using a clustering algorithm to target them with insights about their consumption. The algorithm found four distinct patterns to group as an optimal number and was written in **python**.

**Designed the initial phase of a data pipeline for recovery audit.** I collaborated in the first part by cleaning the data doing the flowchart of feature extraction of all the tables (like ETL process) in **pyspark**. Also I developed the first filter to get the outliers. The client saved time in order to spot outliers from doing manually in excel for several weeks to doing it automated in less than an hour.

### Data science teacher. Mérida México 9/2017- 03/2021

**Syllabus curator. Politechnique University of Yucatan.** Curator of all the path for **data engineering** career, the 20 courses related to big data, computer science, data science and data engineering.

**Teacher of big data (spark and kafka), data preprocessing (python +jupyter), business intelligence (Tableau), data visualization (javascript, D3.js+ processing), data structures (python + C), algorithm analysis, programming paradigms (Scala) and coding principles (python +C) .**

### Business intelligence independent consultant

**Xcaret group** aug 2019. I lectured a seminar about risk mitigation for data leaks, mathematical tools to spot frauds from unknown to quantify them to managers of the amusement park Xcaret and associates. The seminar included exercises in **IPython notebook**. The script helped them to detect anomalies and took seconds instead of a couple of hours in Excel.

## PROJECTS and PUBLICATIONS

- **Scientific paper published in Physica A (ELSEVIER).** “*Wealth condensation in a Barabasi-Albert network*”.
- **Contributed to clean data for analyze a research about chagas disease.** in Yucatán, helping a PhD student at the Research Center for Advanced Studies (CINVESTAV) I created a **R-script** extracted the data from pdf files converting to a json and to map the unstructured json to a structured table in a csv file. . With the script we passed from handling and analyzing only 300 positive cases to over 10k negative cases.

## SKILLS

- **Programming languages:** R, python, scala, javascript, Fortran 77, Latex, Linux shell bash, C, C++, Racket.
- **Programming API and databases:** Jupyter, Apache Spark, RStudio, Mongo DB, SQL, elasticsearch, Google cloud.
- **Math and Stats:** Graduate math tools like Differential equations, Stochastic processes and probability.
- **scientific background:** Critical Thinking, research communication and popular science, teaching for different levels.
- **entrepreneurship background:** business model canvas (social, lean) validation.

## AWARDS and DISTINCTIONS

Jury for startup weekend 2018 Merida Mexico

Jury at HackSureste 2019

third place Startup Weekend Research Edition Feb 2015

first place team of the Hackathon PROFECO beating over four hundred contestants.

first place of physics olympiad for the state team at the national contest (as a trainer).